J. R. Statist. Soc. B (1993)
55, *No.* 1, *pp.* 185–188

Characterizing a Joint Probability Distribution by Conditionals

By ANDREW GELMAN† and T. P. SPEED

University of California, Berkeley, USA

[Received January 1991. Revised October 1991]

SUMMARY

We derive conditions under which a set of conditional and marginal probability distributions will uniquely specify an all-positive joint distribution. Our theoretical result may yield insights into the construction and simulation of multivariate probability models.

Keywords: CONDITIONAL DISTRIBUTIONS; MULTIVARIATE DISTRIBUTIONS

1. INTRODUCTION

In probability modelling and Bayesian statistics, we often encounter multivariate distributions defined by sets of conditional and marginal distributions. It is well known that the set of distributions $\{P(x_i|x_{i+1}, \ldots, x_N), \text{ for } i=1, \ldots, N\}$ uniquely determines the joint distribution $P(x_1, \ldots, x_N)$. It is also possible to specify the joint distribution uniquely, given the set $\{P(x_i|x_j, \text{ all } j \neq i) \text{ for } i=1, \ldots, N\}$ of complete conditional distributions (Besag, 1974; Arnold and Press, 1989). In this latter case, the given information must be checked for consistency; it is possible that no joint probability function exists that is compatible with all the given conditionals.

In contrast, the complete set $\{P(x_i)\}$ of marginal distributions does not come close to specifying the multivariate joint distribution. This paper answers the question: given an *arbitrary* set of conditional and marginal distributions, is their common joint distribution determined uniquely? The ideas here may be applied to the creation and simulation of multivariate probability models such as described by Edwards (1990) and Wermuth and Lauritzen (1990).

To fix ideas, consider the possible ways of defining a joint distribution of two variables (a, b) by using conditional and marginal specifications.

(a) A family of conditional distributions $P(a|b)$ and a marginal distribution $P(b)$ specify the joint distribution uniquely: $P(a, b) = P(a|b)P(b)$, with no restrictions on the two component distributions.

(b) The family of positive conditional distributions $P(a|b)$, along with the single distribution $P(b|a_0)$, for some a_0, uniquely determines the joint distribution:

$$P(a, b) \propto \frac{P(a|b)P(b|a_0)}{P(a_0|b)}.$$

Normalization determines the proportionality constant.

(c) The conditional distributions $P(a|b)$ and $P(b|a)$ determine $P(a, b)$ from the

†*Address for correspondence*: Department of Statistics, University of California at Berkeley, Berkeley, CA 94720, USA.

0035–9246/93/55185 $2.00

above formula for each a_0. The full conditional specification thus over-determines the joint distribution and is self-consistent only if the derived joint distributions agree for all values of a_0.

2. THE PROBLEM

We would like to determine a probability mass function $P(x_1, \ldots, x_N)$, defined with support over a product space of random variables $X_1, \ldots, X_N$. We are thus assuming the positivity condition of Besag (1974). For simplicity we assume that the probability space is discrete; our results also hold for the continuous case, with minor modifications in the proof. We are given a set of marginal and conditional probability mass functions, each of the form $P(A \mid B)$, where the sets A and B are disjoint subsets of $x_1, \ldots, x_N$, and A is not the empty set. For example, if $N=3$, we may be given $\{P(x_1), P(x_2), P(x_1, x_2 \mid x_3), P(x_2 \mid x_1, x_3)\}$. The question is: does a given set of marginal and conditional distributions uniquely specify the joint distribution, assuming that at least one such distribution exists?

We reduce the given information to a set of marginal and conditional probability mass functions, each of the canonical form

$$P(x_i \mid C), \qquad \text{where } C \subset \{x_1, \ldots, x_N\} \setminus \{x_i\}, \tag{1}$$

by replacing each function of the form $P(x_{i_1}, \ldots, x_{i_I} \mid B)$ with the set of I functions: $\{P(x_{i_k} \mid B, (x_{i_l}, \text{all } l > k)), \text{ for } k=1, \ldots, I\}$.

In this notation, probability function (1) is given as a function of x_i and all the variables in C.

3. THEOREM

Consider a collection of marginal and conditional probability specifications of the canonical form (1), consistent with at least one joint distribution with positive support on the product space. The collection uniquely determines the joint distribution if and only if it contains a nested sequence of probability mass functions of the form

$$\{P(y_i \mid A_i, (y_k, \text{all } k > i)), \text{ for } i=1, \ldots, N\}, \tag{2}$$

where $(y_1, \ldots, y_N)$ is a permutation of $(x_1, \ldots, x_N)$ and, for each i, A_i is some subset, possibly empty, of the $i-1$ variables $\{y_1, \ldots, y_{i-1}\}$. Unless all the sets A_i are empty—a hierarchical model—the set of functions (2) must be checked for consistency.

For example, if $N=3$, the joint distribution, if it exists, is uniquely specified by the following set of three conditionals: $\{P(y_1 \mid y_2, y_3), P(y_2 \mid y_3), P(y_3 \mid y_1)\}$.

4. PROOF

4.1. *Step 1*

First assume that a nested set of N conditional distributions is given and is consistent with at least one joint distribution $P(y_1, \ldots, y_N)$; we shall prove by induction that the joint distribution is unique.

(a) $i=1$. $P(y_1 \mid y_2, \ldots, y_n)$ is given by the first distribution in the nested set.

(b) $i=2, \ldots, N$. Assume that $P(y_1, \ldots, y_{i-1} | y_i, \ldots, y_N)$ is specified uniquely by the first $i-1$ distributions in the nested set. The following formula gives the relative values of the conditional probabilities $P(y_1, \ldots, y_i | y_{i+1}, \ldots, y_N)$ for all cases; all probabilities in the formula are implicitly conditional on $y_{i+1}, \ldots, y_N$.

$$\frac{P(y_1, \ldots, y_i)}{P(A_i)} = \frac{P(y_1, \ldots, y_{i-1} | y_i) P(y_i | A_i)}{P(A_i | y_i)}. \tag{3}$$

With any fixed assignment of A_i, we can normalize to obtain the exact probability mass function of $y_1, \ldots, y_i$, conditional on the remaining $N-i$ variables. (If A_i is the empty set, adopt the convention $P(A_i) = P(A_i | y_i) = 1$.)

4.2. *Step 2*

Now assume that the joint distribution is uniquely determined by a set of conditional probability mass functions of the form (1); we shall prove by induction that a nested set of the form (2) must be included.

For each random variable X_i, label two disjoint points as 0 and 1; we define H to be the hypercube product set $\{0, 1\}^N$. This proof uses a set of alternative joint probability mass functions P'_S that differ from P only on the hypercube $H = \{0, 1\}^N$. For any subset S of the random variables $\{X_1, \ldots, X_N\}$, define

$$P'_S(x_1, \ldots, x_N) = P(x_1, \ldots, x_N) \qquad \text{if } (x_1, \ldots, x_N) \text{ is not a vertex of } H.$$

Otherwise,

$$P'_S(x_1, \ldots, x_N) = \begin{cases} P(x_1, \ldots, x_N) - \epsilon & \text{if } \Sigma_{i: X_i \in S} x_i \text{ is even,} \\ P(x_1, \ldots, x_N) + \epsilon & \text{if } \Sigma_{i: X_i \in S} x_i \text{ is odd.} \end{cases}$$

(Since the joint distribution has support on the product space, it is always possible to pick a non-zero ϵ for which the probability mass functions P'_S are all-positive.) Summation over the hypercube H reveals that the conditional probability functions $P'_S(x_k | C)$ and $P(x_k | C)$ differ if and only if $x_k \in S$ and $C \supset S \setminus \{x_k\}$.

(a) $i=1$. Let $S = \{X_1, \ldots, X_N\}$, the complete set of N variables. Then the only conditional distributions in the canonical form (1) that distinguish the probability mass functions P'_S and P are the form $P(y_1 | y_2, \ldots, y_N)$. The first conditional distribution in nested set (2) is thus needed to specify P uniquely.

(b) $i=2, \ldots, N$. Assume that the first $i-1$ distributions of the nested set are given, thus specifying $P(y_1, \ldots, y_{i-1} | y_i, \ldots, y_N)$, and let $S = \{Y_i, \ldots, Y_N\}$. Then the only conditional distributions in the canonical form that distinguish the probability mass functions P'_S and P are those of the form $P(y_k | C)$, where $k \geqslant i$ and $C \supset S \setminus \{y_k\}$. After permuting the labels on Y_i and Y_k, the conditional probability function $P(y_k | C)$ becomes the ith in nested set (2). (Exchanging the labels of Y_i and Y_k does not affect the labelling of the first $i-1$ distributions, which condition on both Y_i and Y_k.)

ACKNOWLEDGEMENTS

We thank a referee for helpful comments and the National Science Foundation for financial support.

REFERENCES

Arnold, B. C. and Press, S. J. (1989) Compatible conditional distributions. *J. Am. Statist. Ass.*, **84**, 152–156.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* B, **36**, 192–225.

Edwards, D. (1990) Hierarchical interaction models. *J. R. Statist. Soc.* B, **52**, 3–20.

Wermuth, N. and Lauritzen, S. L. (1990) On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. R. Statist. Soc.* B, **52**, 21–50.